Un système de segmentation automatique de gestes appliqué à la Langue des Signes

Matilde Gonzalez
IRIT (UPS - CNRS UMR 5505) Université Paul Sabatier,
118 Route de Narbonne,
F-31062 TOULOUSE CEDEX 9
gonzalez@irit.fr

RÉSUMÉ

De nombreuses études sont en cours afin de développer des méthodes de traitement automatique de langues des signes. Plusieurs approches nécessitent de grandes quantités de données segmentées pour l'apprentissage des systèmes de reconnaissance. Nos travaux s'occupent de la segmentation semi-automatique de gestes afin de permettre d'identifier le début et la fin d'un signe dans un énoncé en langue des signes. Nous proposons une méthode de segmentation des gestes à l'aide des caractéristiques de mouvement et de forme de la main.

ABSTRACT

An automatic gesture segmentation system applied to Sign Language

Many researches focus on the study of automatic sign language recognition. Many of them need a large amount of data to train the recognition systems. Our work address the segmentation of gestures in sign language video corpus in order to identify the beginning and the end of signs. We propose an approach to segment gestures using motion and hand shape features.

MOTS-CLÉS : Segmentation de gestes, langue des signes, segmentation de signes.

KEYWORDS: Gesture segmentation, sign language, sign segmentation.

1 Introduction

La langue des signes (LS) est une langue gestuelle développée par les sourds pour communiquer. Un énoncé en LS consiste en une séquence de signes réalisés par les mains, accompagnés d'expressions du visage et de mouvements du haut du corps, permettant de transmettre des informations en parallèles dans le discours. Même si les signes sont définis dans des dictionnaires, on trouve une très grande variabilité liée au contexte lors de leur réalisation. De plus, les signes sont souvent séparés par des mouvements de co-articulation (aussi appelé *'transition'*). Un exemple est montré dans la Figure 1. Cette extrême variabilité et l'effet de co-articulation représentent un problème important dans la segmentation automatique de gestes.

Une méthode permettant de segmenter semi-automatiquement des énoncés en LS, sans utiliser d'apprentissage automatique est présenté. Plus précisément, nous cherchons à détecter les limites de début et fin de signes. Cette méthode de segmentation de gestes nécessite plusieurs traitements de bas niveau afin d'extraire les caractéristiques de mouvement et de forme de la main. Les

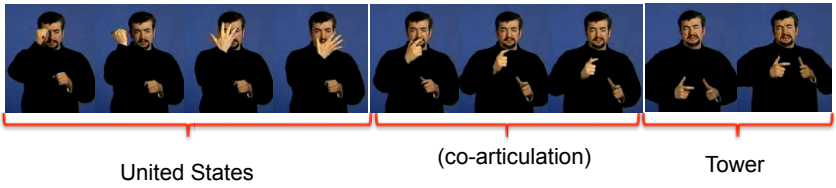

FIGURE 1 – Exemple de *co-articulation* : geste netre la fin du signe "Etats-Unis" et le debout du signe "tour" en Langue de Signes Française.

caractéristiques de mouvement sont utilisées pour réaliser une première segmentation qui est par la suite améliorée grâce à l'utilisation de caractéristiques de forme. En effet,celles-ci permettent de supprimer les limites de segmentation détectées en milieu des signes.

Cet article est structuré comme suit. La section 2 présente une synthèse des méthodes de segmentation automatique appliquées à la LS. Nous montrons ensuite dans la section 3 l'extraction de caractéristiques de mouvement et de forme afin de segmenter les gestes dans la séquence vidéo. Des résultats expérimentaux sont ensuite présentés en section 4. Enfin, en section 5, une conclusion rappelle les principaux résultats obtenus et évoque quelques perspectives de recherche.

2 Segmentation Automatique des Signes : état de l'art

Actuellement plusieurs recherches s'intéressent au problème de l'analyse automatique de la LS (Ong et Ranganath, 2005), plus particulièrement de sa reconnaissance (Imagawa *et al.*, 1998; Starner et Pentland, 1995; Zieren *et al.*, 2006). Dans (Grobel et Assan, 1997) les données d'apprentissage sont des signes isolés réalisés plusieurs fois par un ou plusieurs signeurs. La réalisation des signes est dépendante du contexte et, dans le cas des signes isolés, la co-articulation n'est pas prise en compte. En ce qui concerne la segmentation automatique de gestes en LS, Nayak *et al.* (Nayak *et al.*, 2009) ont proposé une méthode qui permet d'extraire automatiquement les limites d'un signe à l'aide de plusieurs occurrences du signe dans la vidéo. Ils considèrent la forme et la position relative des mains par rapport au corps. Pour la plupart des signes ces caractéristiques varient énormément selon le contexte cantonnant cette approche à quelques exemples typiques. Lefebvre et Dalle (Lefebvre-Albaret et Dalle, 2010) ont présenté une méthode utilisant des caractéristiques de bas niveau afin de segmenter semi-automatiquement les signes. Ils ne considèrent que le mouvement dans le but d'identifier plusieurs types de symétries. Or plusieurs signes sont composés de plusieurs séquences avec différents types de symétrie, ces signes seront sur-segmentés.

Afin de résoudre certains problèmes émergents de l'état de l'art nous proposons une méthode de segmentation automatique des signes qui exploite le caractéristiques de mouvement, et de forme de la main.

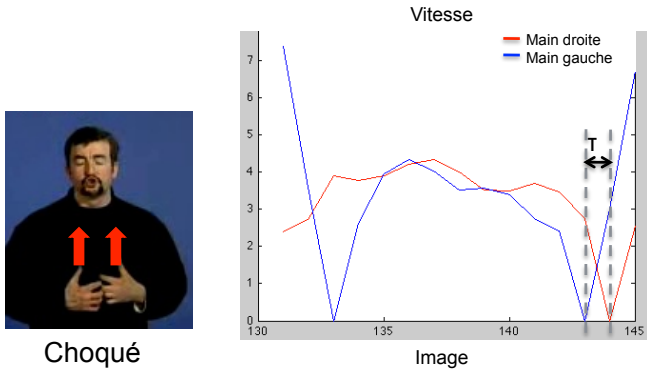

FIGURE 2 – Signe "choqué" en LSF et vitesse des deux mains.

3 Segmentation automatique de gestes

La segmentation des signes correspond à la détection du début et de la fin d'un signe. Pour cela nous utilisons les résultats de suivi de composantes corporelles (Gonzalez et Collet, 2011) afin de segmenter les signes grâce à des caractéristiques de mouvement. Ensuite la forme de la main est utilisée pour améliorer les résultats de segmentation (Gonzalez et Collet, 2010).

3.1 Classification du mouvement

Les caractéristiques de mouvement sont extraites à partir des résultats du suivi des composantes corporelles. Les vitesses des mains droite et gauche, $v_1(t)$ et $v_2(t)$ sont calculées à l'aide des positions des mains pour chaque image. La norme de la vitesse est utilisée pour le calcul de la vitesse relative $v_r(t)$, c'est-à-dire la différence entre la vitesse de la main gauche et celle de la main droite. Quand les mains bougent ensemble nous remarquons un léger décalage entre les profils de vitesses des deux mains bien que leur allure reste très proche comme on peut le voir avec le signe "Choqué" (Fig. 2).

Grâce à la vitesse relative nous déterminons les séquences statiques, aucune main ne bouge, ou celles réalisées avec une ou deux mains. A partir de cette classification nous pouvons identifier les événements définis comme les début et fin potentiels de signes et détectés comme un changement de classe. Toutefois cette approche détecte des événements en milieu de signe. On dit alors que les séquences ont été sur-segmentées. Par exemple la figure 3(gauche) illustre la réalisation du signe "*Quoi ?*" en LSF. Il s'agit d'un signe symétrique répété où les deux mains bougent simultanément en direction opposée. La figure 3(droite) montre les évènements détectés en fonction des classes définies précédemment. La segmentation peut être améliorée en tenant compte de la forme des mains car, pour ce signe comme pour beaucoup d'autres, la configuration des mains reste inchangée.

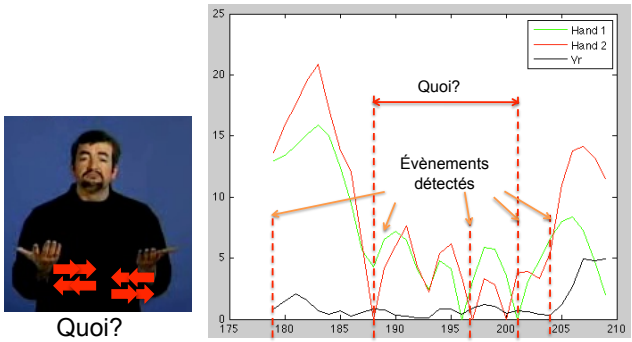

FIGURE 3 – Signe *'Quoi ?'* en LSF et les vitesses pour les deux mains, la vitesse relative et les événements détectés.

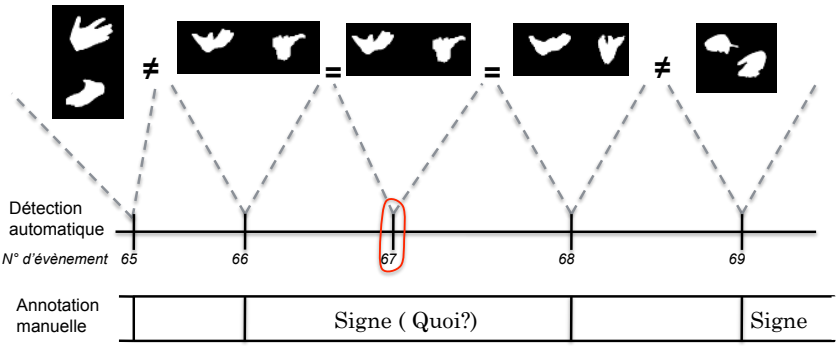

FIGURE 4 – Illustre les mains segmentées pour chaque événement détecté ainsi que la vérité-terrain.

3.2 Caractérisation de la forme des mains

Dans cette étape nous introduisons des informations sur la forme de la main afin de corriger la sur-segmentation. La reconnaissance de la configuration de la main est un problème complexe du fait de la grande variabilité de la forme 2D obtenue à l'aide d'une seule caméra.

Afin d'extraire les caractéristiques de forme, nous devons d'abord segmenter les mains pour chaque événement. La forme de la main est systématiquement comparée avec celle des événements adjacents. Nous utilisons deux mesures de similarité : le diamètre équivalent ϵ_d et l'excentricité ϵ. L'avantage d'utiliser ces types de mesures est l'invariance en translation et en rotation. Cependant l'inconvénient est la sensibilité au changement d'échelle et au bruit. La figure 4 montre les résultats de segmentation du signe "*Quoi ?*" en LSF. L'étape précédente a segmenté le signe en tenant compte des caractéristiques de mouvement ce qui a entrainé la sur-segmentation du signe. Nous remarquons que la forme des mains reste similaire entre certains événements détectés. On supprime donc celui du milieu pour corriger la segmentation.

4 Résultats expérimentaux

Nous avons réalisé l'évaluation à l'aide de deux séquences vidéo sans aucune contrainte sur la langue : LS Colin et DEGELS. L'algorithme de segmentation a été appliqué sur 2500 images. Les vérités-terrain pour les deux séquences ont été manuellement réalisées par un signeur sourd-né. L'évaluation consiste à à compter les événements correctement segmentés en tenant compte d'une tolérance (TP : true positifs) et les événements détectés mais qui ne correspondent pas à une limite annotée (FP : False positif). La tolérance δ pour le calcul du taux de TP a été déterminée expérimentalement. Un signeur expérimenté a annoté une séquence vidéo plusieurs fois afin de déterminer sa variabilité qui s'élève dans notre cas à 1,7 images en moyenne. La segmentation est considérée comme correcte si le nombre d'images entre l'annotation et l'événement détecté est proche à la variabilité du signeur. Le tableau 4 montre les résultats pour les deux séquences vidéo avec une tolérance de deux images. On remarque qu'à l'introduction des caractéristiques de forme de la main le taux de TP reste le même alors que le taux de FP diminue de 3% pour LS-Colin et de 10% pour le corpus Degels.

	Motion		Motion + Hand Shape	
	TP(%)	FP(%)	TP(%)	FP(%)
LS- Colin	81.6	46.2	81.6	44.9
DEGELS	74.5	54.2	74.5	44.7

TABLE 1 – Résultats de segmentation de gestes

5 Conclusion

Nous présentons ici un système de segmentation temporelle de séquences vidéo en LS. La segmentation a été réalisée en ne considérant que des caractéristiques de bas niveau, ce qui rend notre méthode généralisable pour toutes les LS. Nous utilisons d'abord les caractéristiques de mouvement extraites à l'aide de notre algorithme de suivi qui est robuste aux occultations. Ensuite grâce aux caractéristiques de forme de la main nous sommes capable de corriger la segmentation. Cette méthode a montré des résultats prometteurs qui peuvent être utilisés pour la reconnaissance de signes et pour l'annotation en gloses des séquences à l'aide d'un modèle linguistique de la LS.

Remerciements

Ces recherches sont financées par le 7ème programme cadre Communauté Européenne (FP7/2007-2013) accord no 231135.

Références

Gonzalez, M. et Collet, C. (2010). Head tracking and hand segmentation during hand over face occlusion in sign language. *In Int. Workshop on Sign, Gesture, and Activity (ECCV)*.

Gonzalez, M. et Collet, C. (2011). Robust body parts tracking using particle filter and dynamic template. *In 18th IEEE ICIP*, pages 537–540.

Grobel, K. et Assan, M. (1997). Isolated sign language recognition using hidden markov models. *In IEEE Int. Conference on Systems, Man, and Cybernetics*, volume 1, pages 162–167. IEEE.

Imagawa, K., Lu, S. et Igi, S. (1998). Color-based hands tracking system for sign language recognition. *In Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 462–467.

Lefebvre-Albaret, F. et Dalle, P. (2010). Body posture estimation in sign language videos. *Gesture in Embodied Communication and HCI*, pages 289–300.

Nayak, S., Sarkar, S. et Loeding, B. (2009). Automated extraction of signs from continuous sign language sentences using iterated conditional modes. *CVPR*, pages 2583–2590.

Ong, S. et Ranganath, S. (2005). Automatic sign language analysis : A survey and the future beyond lexical meaning. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, pages 873–891.

Starner, T. et Pentland, A. (1995). Real-time american sign language recognition from video using hidden markov models. *In Proc. International Symposium on Computer Vision*, pages 265–270.

Zieren, J., Canzler, U., Bauer, B. et Kraiss, K. (2006). Sign language recognition. *Advanced Man-Machine Interaction*, pages 95–139.